

## Durham Research Online

---

### Deposited in DRO:

13 September 2018

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Morlon, Hélène and Lewitus, Eric and Condamine, Fabien L. and Manceau, Marc and Clavel, Julien and Drury, Jonathan (2016) 'RPANDA : an R package for macroevolutionary analyses on phylogenetic trees.', *Methods in ecology and evolution.*, 7 (5). pp. 589-597.

### Further information on publisher's website:

<https://doi.org/10.1111/2041-210X.12526>

### Publisher's copyright statement:

This is the accepted version of the following article: Morlon, Hélène, Lewitus, Eric, Condamine, Fabien L., Manceau, Marc, Clavel, Julien Drury, Jonathan (2016). RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods in Ecology and Evolution* 7(5): 589-597, which has been published in final form at <https://doi.org/10.1111/2041-210X.12526>. This article may be used for non-commercial purposes in accordance With Wiley Terms and Conditions for self-archiving.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

**RPANDA: an R package for macroevolutionary analyses on  
phylogenetic trees**

Morlon, H.<sup>1\*</sup>, Lewitus, E.<sup>1</sup>, Condamine, F.L.<sup>2</sup>, Manceau, M.<sup>1</sup>, Clavel, J.<sup>1</sup>, Drury, J.<sup>1</sup>

<sup>1</sup> CNRS UMR 8197, Institut de Biologie, Ecole Normale Supérieure, 46 rue d'Ulm, 75005  
Paris

<sup>2</sup> CNRS UMR 5554, Institut des Sciences de l'Evolution, Montpellier, Place Eugène  
Bataillon, 34000 Montpellier, France

\*Corresponding author: morlon@biologie.ens.fr

**Short title:** RPANDA package for macroevolutionary analyses

## Summary

1. A number of approaches for studying macroevolution using phylogenetic trees have been developed in the last few years. Here, we present RPANDA, an R package that implements model-free and model-based phylogenetic comparative methods for macroevolutionary analyses.
2. The model-free approaches implemented in RPANDA are recently developed approaches stemming from graph theory that allow summarizing the information contained in phylogenetic trees, computing distances between trees, and clustering them accordingly. They also allow identifying distinct branching patterns within single trees.
3. RPANDA also implements likelihood-based models for fitting various diversification models to phylogenetic trees. It includes birth-death models with i) constant, ii) time-dependent, and iii) environmental-dependent speciation and extinction rates. It also includes models with equilibrium diversity derived from the coalescent process, as well as a likelihood-based inference framework to fit the individual-based model of Speciation by Genetic Differentiation, which is an extension of Hubbell's Neutral Theory of Biodiversity.
4. RPANDA can be used to: i) characterise trees by plotting their spectral density profiles ii) compare trees and cluster them according to their similarities, iii) identify and plot distinct branching patterns within trees, iv) compare the fit of alternative diversification models to phylogenetic trees, ii) estimate rates of speciation and extinction, iii) estimate and plot how these rates have varied with time and environmental variables, and iv) deduce and plot estimates of species richness through geological time.

5. RPANDA provides investigators with a set of tools for exploring patterns in phylogenetic trees and fitting various models to these trees, thereby contributing to the on-going development of phylogenetics in the life sciences.

**Keywords:** diversification rates, speciation, extinction, likelihood, macroevolution, speciation by genetic differentiation, graph Laplacian, spectral density profiles.

## Introduction

Phylogenetic approaches have become a central component of various areas of the life sciences. A number of packages are available to handle and utilise phylogenetic trees (e.g. ‘ape’ Paradis *et al.* 2004), in order to understand, for example, community assembly (e.g. ‘picante’ Kembel *et al.* 2010, ‘DAMOCLES’ Pigot & Etienne 2015), trait evolution (e.g. ‘Coevol’ Lartillot & Poujol 2011, ‘geiger’ Pennell *et al.* 2014), and diversification (e.g. ‘BayesRate’ Silvestro *et al.* 2011, ‘TreePar’ Stadler 2011a, ‘diversitree’ Fitzjohn 2012, ‘DDD’ Etienne *et al.* 2012, ‘geiger’ Pennell *et al.* 2014, ‘BAMM’ Rabosky *et al.* 2014).

Here, we present and describe the R package RPANDA, which implements both model-free and model-based phylogenetic approaches that are not implemented in previous packages. The model-free approaches are inspired from graph theory and described in detail in Lewitus & Morlon (2015). They are designed to compare and classify phylogenetic trees without any *a priori* formulation of a model of cladogenesis underlying tree shape. The model-based approaches aim to fit various models of cladogenesis to phylogenetic trees by maximum likelihood. They are described in detail in Morlon *et al.* (2010, 2011), Condamine *et al.* (2013) and Manceau *et al.* (2015).

Within the category of model-based approaches, the package includes three main classes of models: birth-death models, coalescent models, and individual-based models (see Morlon 2014 for a review of these three different types of models). Birth-death models are those that were originally considered by Nee *et al.* (1992) and are at the basis of most diversification models used today. The present package allows fitting of birth-death models: i) with speciation and extinction rates varying as a function of time, with any type of functional form and with potential periods of diversity decline (Morlon *et al.* 2011), and ii) with speciation and extinction rates varying as a function of any variable that has been measured through geological time (e.g., temperature, Condamine

*et al.* 2013). Models based on the coalescent process were designed to consider scenarios of equilibrium diversity (Morlon *et al.* 2010). Finally, individual-based models have been extensively studied in ecology, but likelihood-based inferences from phylogenies are typically not available. The present package allows likelihood inference for the model of Speciation by Genetic Differentiation (Manceau *et al.* 2015), which is an extension of the Neutral Theory of Biodiversity (Hubbell 2001).

## **Description**

RPANDA is an R package (R Development Core Team 2014) that can be installed from the CRAN repository (<http://cran.r-project.org>). RPANDA relies on the R packages ‘ape’ (Paradis *et al.* 2004), ‘picante’ (Kembel *et al.* 2010), ‘phytools’ (Revell 2012), ‘deSolve’ (Soetaert *et al.* 2010), and ‘igraph’ (Csardi & Nepusz 2006).

The main functions of the package are listed in Table 1, classified into functions related to the model-free and model-based approaches. Nearly every function is associated with a ‘plot’ function that helps visualize the results of the analyses.

To illustrate the use of RPANDA, we analyse the phylogeny of the bat family Phyllostomidae. This phylogeny is the maximum clade credibility tree used in Rolland *et al.* (2014), which originally comes from the mammalian supertree (Bininda-Emonds *et al.* 2007; Fritz *et al.* 2009); it contains 150 of the 165 known bat species (i.e., it is 91% complete). To begin, we open an R console, and we install and load the RPANDA package as well as the example datasets.

```
> install.packages('RPANDA', dependencies=TRUE)
> library(RPANDA)
> data(Phyllostomidae)
> data(Phyllostomidae_genera)
```

'Phyllostomidae' is the family-level phylogeny and 'Phyllostomidae\_genera' is a list of 25 phylogenies corresponding to Phyllostomidae genera with more than one species.

### **Characterising and comparing phylogenies using spectral densities**

We recently developed a new approach, described in detail in Lewitus & Morlon (2015), to efficiently summarize the shape of a phylogenetic tree. This approach can be used to measure similarities between trees and to cluster them accordingly, for example in order to identify phylogenies shaped by similar versus dissimilar diversification patterns. It can also help in identifying regions of a tree that have distinct branching patterns, which can for example reflect shifts in modes or rates of diversification. We summarize the shape of a phylogeny by its spectral density, which is a smoothed version of the frequencies of eigenvalues associated with a matrix (the graph Laplacian) built from the pairwise phylogenetic distances between nodes (see Lewitus & Morlon 2015 for a more detailed description). The function `spectR` computes the eigenvalues associated with a given phylogeny, and characteristics associated with the spectrum of eigenvalues, namely the principal eigenvalue, asymmetry (skewness), two measures of peakedness (kurtosis and peak height), and eigengap. The eigengap is given by the position of the largest difference between successive eigenvalues listed in descending order. This number is related to the number of peaks in the spectral density plot and is indicative of the number of modalities (i.e. distinct branching patterns) in a phylogeny.

```
> res<-spectR(Phyllostomidae)
```

returns the above information for the Phyllostomidae phylogeny. In particular, `res$eigengap` returns the number of modalities, suggesting three distinct branching patterns in this bat family.

```
> plot_spectR(res)
```

displays the spectral density profile and a plot of the eigenvalues ranked in descending order (Figure 1).

Once the putative number of modalities is identified, the `BICompare` function can be used to assess the significance of these modalities and to identify their location on the phylogeny. The statistical significance of the modalities is assessed by comparing the Bayesian Information Criterion (BIC) for detecting  $i$  clusters in the distance matrix of the empirical phylogeny and in randomly bifurcating trees parameterized on that tree (Lewitus & Morlon 2015). The function also identifies the location of the distinct branching patterns on the phylogeny by k-means clustering and returns the ratio of between-cluster sum of squares (BSS) to total sum of squares (TSS) for the clustering identified by the algorithm. The highest the BSS/TSS ratio, the more distinct the modalities are from each other. Different iterations of the k-means clustering algorithm can lead to different modality configurations, and BSS/TSS values allow the comparison between these configurations (configurations with high BSS/TSS should be preferred).

```
> res<-BICompare(Phyllostomidae,3)
```

returns the above information for the Phyllostomidae phylogeny. The BIC score for the Phyllostomidae phylogeny is nearly a magnitude smaller than it is for the randomly bifurcating trees parameterized on that phylogeny, suggesting that the three modalities are significant. Typically, a BIC ratio  $\leq 0.25$  is deemed significant.

```
> plot_BICompare(Phyllostomidae,res)
```

displays the Phyllostomidae phylogeny with branches coloured according to the modality they belong to, as assessed by the k-means clustering algorithm (Figure 2).

Spectral densities are particularly useful for comparing phylogenies. The `JSDtree` function computes the pairwise distances between a list of phylogenies, measured as the Jensen-Shannon distance between their spectral densities. The



150 JSDtree\_cluster function uses these pairwise distances (or potentially other distance  
151 metrics) to cluster phylogenies into groups. The clustering is implemented using the  
152 hierarchical clustering and k-medoid clustering algorithms. To illustrate this approach,  
153 we compare the 25 phylogenies corresponding to Phyllostomidae genera with more  
154 than one species.

```
155 > res<-JSDtree(Phyllostomidae_genera)
```

156 returns the matrix containing the pairwise Jensen-Shannon distances between the 25  
157 phylogenies.

```
158 > JSDtree_cluster(res)
```

159 plots the heatmap and hierarchical cluster (Figure 3) as well as the hierarchical cluster  
160 with bootstrap support (not shown here). It also returns the optimal number of clusters  
161 given by the k-medoids algorithm, here suggesting that Phyllostomidae genera cluster  
162 into two meaningful groups. The function returns the assignment of each phylogeny to  
163 each of the two groups, as well as a measure of statistical support for this assignment.

164

### 165 **Fitting models of diversification to phylogenies**

166 One of the most popular approaches for analysing the diversification of clades consists  
167 in fitting various models of diversification to molecular phylogenies using maximum  
168 likelihood inference, comparing the likelihood support of the different models, and  
169 estimating the parameters of the model (see Morlon 2014 for a review). The different  
170 types of functions available in RPANDA reflect this general approach (Table 1): the  
171 ‘likelihood’ functions compute the likelihood associated with different  
172 diversification models, the ‘fit’ functions fit the corresponding models by maximum  
173 likelihood, and the ‘plot’ functions plot estimates of how various variables (e.g.  
174 speciation and extinction rates, species richness) have varied as a function of time or

various environmental factors (e.g. temperature). Simulating phylogenies under the different models is often useful, for example, to test the power of the approach to recover true parameter values or to measure type I & II error rates. We do not generally provide functions to simulate phylogenies, as they are available in other packages such as TreeSim (Stadler 2011b) and TESS (Höhna 2013). The one exception is the `sim_sgd` function that simulates phylogenies under the model of Speciation by Genetic Differentiation (Manceau *et al.* 2015).

Fitting a diversification model to a phylogeny consists in finding the parameters that maximize the likelihood associated with the model. The ‘`fit`’ functions, therefore, take as argument, at minimum, a phylogeny (`phylo`), initial parameter values (`par`), and the maximization algorithm to be used (`meth`). The various likelihood expressions all depend on the fraction of extant species that are sampled in the phylogeny; this fraction (`f`), therefore, also needs to be specified. Finally, `tot_time` specifies the age of the phylogeny. We often have access only to the crown age, in which case `tot_time` is given by `max(node.age(phylo)$ages)`. If the stem age is known, specifying this older age in `tot_time` can provide a different and complementary inference.

### *Time-dependent diversification models*

RPANDA can be used to test whether (and how) diversification rates varied through time (Morlon *et al.* 2011). To illustrate the approach we fit a birth-death model with time-varying rates to the Phyllostomidae phylogeny. We first need to specify the assumed functional form of the time-dependency. For example, if we want to fit a model with an exponential variation of the speciation rate with time (`f.lamb`), and a constant extinction rate (`f.mu`), we define these functions as follows:

```

199 > f.lamb<-function(t,y){y[1]*exp(y[2]*t)}
200 > f.mu<-function(t,y){y[1]}
201 For a linear dependency of the speciation rate, the function would be:
202 > f.lamb.lin<-function(t,y){y[1]+y[2]*t}
203 The variable t represents time, running from the present to the past, while the variable
204 y is a vector containing the different parameters involved in the definition of the
205 temporal dependency. The parameters in y are therefore the parameters that will be
206 estimated by maximum likelihood. We need to specify initial values for these
207 parameters, for example
208 > lamb_par_init<-c(0.05,0.01)
209 sets the initial parameter values defining the f.lamb speciation function, and
210 > mu_par_init<-c(0.005)
211 sets the initial parameter values defining the f.mu extinction function. The result of the
212 fitting procedure should not depend on the choice of the initial parameter values, which
213 can be checked by running the model with several sets of (realistic) initial values. For
214 example, the speciation rate at present (lamb_par_init[1]) typically takes value
215 ranging between 0.01 and 1 (event per lineage, per million years). The rate of variation
216 of the speciation rate (lamb_par_init[2]) can then be chosen such that the resulting
217 speciation rate (output of f.lamb(t,lamb_par_init)) remains within this realistic
218 range throughout the clade history (i.e. for t ranging from 0 to crown or stem age). And
219 finally, the initial extinction parameters set in mu_par_init can be chosen such that the
220 resulting extinction rate (output of f.mu(t,mu_par_init)) is smaller than the
221 speciation rate at the beginning of clade's history (i.e. for t set to the crown or stem age)

```

222 and remains positive throughout the clade history (i.e. for  $t$  ranging from 0 to crown or  
223 stem age).

224 We can now fit the model. If we knew the stem age of Phyllostomidae, we could specify  
225 this age in `tot_time`. As we do not have this information, we define:

```
226 > tot_time<-max(node.age(Phyllostomidae)$ages)
```

227 Finally, we fit the model by maximum likelihood using the following command:

```
228 > res<-fit_bd(Phyllostomidae,tot_time,f.lamb,f.mu,lamb_par_init,  
229               mu_par_init,f=150/165,expo.lamb=TRUE,cst.mu=TRUE)
```

230 The two options `expo.lamb=TRUE` and `cst.mu=TRUE` are set to `TRUE` to speed up the  
231 computation by using analytical solutions; such solutions have been implemented for  
232 exponential, linear, and constant functions. If the options are set to `TRUE` when the  
233 time-dependency is not of the proper form (exponential, linear, or constant), the code  
234 will not return the proper answer. When in doubt, it is better to prefer the default `FALSE`  
235 option. There is an additional `dt` option (not used in the examples here) that can also  
236 speed up the computation by using piecewise constant approximations in the  
237 computation of the integrals. There is also an option that specifies whether we are  
238 working with crown or stem ages, which has consequences for the conditioning that  
239 should be used in the computation of the likelihood: the process should be conditioned  
240 on survival when working with stem ages, and conditioned on a speciation event at the  
241 crown and survival of the two descending lines when working with crown ages (Morlon  
242 *et al.* 2011). The default (used here) is the `cond="crown"` option, but it should be set to  
243 `cond="stem"` if `tot_time` is the stem age. The output `res` of the fit contains the  
244 maximum log-likelihood value (-469.36), the corrected Akaike information criterion  
245 (AICc) (944.89), and the maximum likelihood parameter estimates. For example,

```

246 > res$lamb_par[1]
247 returns the maximum parameter estimate of  $y[1]$ , which is the speciation rate at time
248  $t=0$ , i.e. the present (here 0.099).
249 > res$lamb_par[2]
250 returns the maximum parameter estimate of  $y[2]$ , which is the rate of change in
251 speciation rate, with time running from the present to the past (here 0.022). A positive
252 rate of change with time running from the present to the past – as estimated here –
253 suggests a negative rate of change (decline in speciation rate) during the clade's history.
254 > plot_fit_bd(res,tot_time)
255 returns three plots, which represent speciation, extinction and net diversification (i.e.
256 speciation minus extinction) rates through time (Figure 4). If a model without extinction
257 is fitted:
258 > f.mu<-function(t,y){0}
259 > mu_par_init<-c()
260 > res_noext<-fit_bd(Phyllostomidae,tot_time,f.lamb,f.mu,
261 lamb_par_init,mu_par_init,f=150/165,expo.lamb=TRUE,fix.mu=TRUE)
262 > plot_fit_bd(res_noext,tot_time)
263 returns two plots, which represent speciation and net diversification rates through time
264 (in this case, these two rates are equal).
265 Once estimates of the temporal variation in speciation and extinction rates have been
266 obtained (as described above), estimates of how species richness varied through time
267 can be computed by resolving the appropriate differential equation (Morlon et al. 2011;
268 Morlon 2014).
269 > plot_dtt(res,tot_time,N0=165)

```

270 plots the result of this estimation procedure, as illustrated in Figure 5.

271

272 *Environmental-dependent diversification models*

273 RPANDA can also be used to test the potential effect that past environmental conditions  
274 had on diversification (Condamine *et al.* 2013). Fitting the environmental-dependent  
275 birth-death model is very similar to fitting the time-dependent birth-death model. In  
276 addition to a phylogeny, this model requires knowledge (typically an estimate) of how a  
277 given environmental variable varied through time. The example provided in RPANDA is  
278 temperature variation through the Cenozoic, estimated using oxygen isotope ratios  
279 (Zachos *et al.* 2008). We begin by loading the temperature data:

280 `> data(InfTemp)`

281 `InfTemp` is a two-column dataframe in which the first column reports time (measured  
282 from the present to the past) and the second column reports the corresponding  
283 estimated temperature at each time (Condamine *et al.* 2013). Any other (abiotic or  
284 biotic) environmental variable in this format can be used in place of `InfTemp` in order to  
285 test the potential effect that this variable had on diversification (e.g. sea-level  
286 fluctuations, Condamine *et al.* 2015). Here, diversification rates can depend on time as  
287 well as on the environmental variable; the `f.lamb` and `f.mu` functions therefore take  
288 two arguments (time `t` and the environmental variable `x`) in addition to the parameters  
289 `y` to be estimated. We fit a simple model with an exponential dependence of the  
290 speciation rate on the environmental variable, no time dependence, and no extinction.  
291 We thus define:

292 `> f.lamb<-function(t,x,y){y[1]*exp(y[2]*x)}`

293 and

```

294 > f.mu<-function(t,x,y){0}
295 as well as initial parameter estimates:
296 > lamb_par_init<-c(0.10,0.01)
297 > mu_par_init<- c()
298 Finally, we fit the model:
299 > res<-fit_env(Phyllostomidae,InfTemp,tot_time,f.lamb,f.mu,
300 lamb_par_init,mu_par_init,f=150/165,fix.mu=TRUE,dt=1e-3)
301 Note that we do not use the option expo.lamb=TRUE, as f.lamb is an exponential
302 function of the environmental variable, not an exponential function of time. Setting
303 expo.lamb to TRUE would yield spurious results. However, we can speed up the
304 computation by specifying dt=1e-3, which uses a piece-wise constant approximation in
305 the evaluation of integrals. The output res of the fit contains the maximum log-
306 likelihood value (-468.44), the AICc (940.97), and the maximum likelihood parameter
307 estimates. For example,
308 > res$lamb_par[1]
309 returns the maximum parameter estimate of y[1], which is the speciation rate that
310 would correspond to a temperature of 0°C (here 0.077).
311 > res$lamb_par[2]
312 returns the maximum parameter estimate of y[2], which is the rate of change in
313 speciation rate with temperature (here 0.083). A positive value thus suggests a
314 positive effect of the environmental variable (here temperature) on speciation rates.
315 Note that the environmental model fitted here receives a better support than the model
316 with an exponential dependency of speciation rate with time ( $\Delta AICc = 944.89 - 940.97 =$ 
317 3.92).

```

```

318 > plot_fit_env(res, InfTemp, tot_time)
319 plots diversification rates as a function of the environmental variable and time (Figure
320 6).
321
322 The model of Speciation by Genetic Differentiation
323 RPANDA also contains functions to fit a modified version of Hubbell's neutral model of
324 biodiversity (Hubbell 2001): the model of Speciation by Genetic Differentiation, which is
325 described in detail in Manceau et al. 2015. We can fit the model following a similar
326 procedure. The model is individual-based, and involves parameters describing
327 the birth and death of individuals, as well as a per-individual mutation rate. We define
328 initial parameter values for the birth, growth (birth minus death), and mutation rates
329 (given in events per Myr if the branch lengths of the phylogeny are measured in Myrs),
330 for example:
331 > par_init<-c(1e7, 1e7-0.5, 0.8)
332 We can then fit the model (this takes more time than the previous examples, as
333 computing likelihoods requires integrating a set of coupled differential equations along
334 the tree, Manceau et al. 2015) :
335 > fit_sgd(Phyllostomidae, tot_time, par_init, f=150/165)
336 returns the maximum log-likelihood (-466), the AICc (938), and the estimated birth
337 (1e7), growth (0.157), and mutation (0.198) rates. As explained in Manceau et al.
338 (2015), the likelihood surface is quite flat with respect to the birth rate, such that this
339 parameter cannot be estimated with confidence. Birth rate estimates are thus sensitive
340 to the choice of the initial parameter values and should not be trusted.
341

```



## Resources

RPANDA is an open source package available for download on the CRAN repository at <https://cran.r-project.org/web/packages/RPANDA/index.html>. It includes illustrative data and a detailed manual. The package is constantly evolving. Bayesian implementations and functions that allow fitting new models of phenotypic evolution, such as the matching competition model (Drury *et al.* 2015), are already under development. Contributions are welcome; automatic tests have been implemented to facilitate a collaborative development and to insure the replicability of results. The most recent version is available on github at <https://github.com/hmorlon/PANDA>.

## Acknowledgement

We thank members of HM's group, as well as reviewers D.S. Caetano and D. Silvestro for comments on a previous version of the paper. The development of RPANDA is supported by grant CHEX-ECOEVOBIO from the Agence Nationale de la Recherche and grant 616419-PANDA from the European Research Council awarded to HM.

## Data Accessibility

All data used in this manuscript is available through the R package RPANDA, available on CRAN <https://cran.r-project.org/web/packages/RPANDA/index.html>.

## References

Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L. & Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, **446**, 507–512.

Condamine, F.L., Rolland, J. & Morlon, H. (2013). Macroevolutionary perspectives to

367 environmental change. *Ecology Letters*, **16**, 72–85.

368 Condamine, F.L., Toussaint, E.F.A., Clamens, A.-L., Genson, G., Sperling, F.A.H. & Kergoat,  
369 G.J. (2015). Deciphering the evolution of birdwing butterflies 150 years after Alfred  
370 Russel Wallace. *Scientific Reports*, **5**.

371 Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network  
372 research. *InterJournal, Complex Systems*, **1695**, 1–9.

373 Drury, J., Clavel, J. & Morlon, H. (2015). Estimating the effect of competition on trait  
374 evolution using maximum likelihood. *bioRxiv*.

375 Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A. & Phillimore, A.B.  
376 (2012). Diversity-dependence brings molecular phylogenies closer to agreement with  
377 the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1300–1309.

378 FitzJohn, R.G. (2012). Diversitree: comparative phylogenetic analyses of diversification  
379 in R. *Methods in Ecology and Evolution*, **3**, 1084–1092.

380 Fritz, S.A., Bininda-Emonds, O.R.P. & Purvis, A. (2009). Geographical variation in  
381 predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology*  
382 *Letters*, **12**, 538–549.

383 Höhna, S. (2013). Fast simulation of reconstructed phylogenies under global time-  
384 dependent birth–death processes. *Bioinformatics*, **29**, 1367–1374.

385 Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*.  
386 Princeton University Press, Princeton.

387 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D.,  
388 Blomberg, S.P. & Webb, C.O. (2010). Picante: R tools for integrating phylogenies and  
389 ecology. *Bioinformatics*, **26**, 1463–1464.

390 Lartillot, N. & Poujol, R. (2011) A phylogenetic model for investigating correlated  
391 evolution of substitution rates and continuous phenotypic characters. *Molecular Biology*

392 *and Evolution*, **28**, 729-744.

393 Lewitus, E. & Morlon, H. (2015). Characterizing and comparing phylogenies from their  
 394 Laplacian spectrum. *bioRxiv*.

395 Manceau, M., Lambert, A. & Morlon, H. (2015). Phylogenies support out-of-equilibrium  
 396 models of biodiversity. *Ecology letters*, **18**, 347–356.

397 Morlon, H. (2014). Phylogenetic approaches for studying diversification. *Ecology Letters*,  
 398 **17**, 508–525.

399 Morlon, H., Parsons, T.L. & Plotkin, J.B. (2011). Reconciling molecular phylogenies with  
 400 the fossil record. *Proceedings of the National Academy of Sciences of the United States of*  
 401 *America*, **108**, 16327–16332.

402 Morlon, H., Potts, M.D. & Plotkin, J.B. (2010). Inferring the dynamics of diversification: a  
 403 coalescent approach. *PLoS Biology*, **8**, e1000493.

404 Nee, S., Mooers, A.O. & Harvey, P.H. (1992). Tempo and mode of evolution revealed from  
 405 molecular phylogenies. *Proceedings of the National Academy of Sciences*, **89**, 8322–8326.

406 Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of Phylogenetics and  
 407 Evolution in R language. *Bioinformatics*, **20**, 289–290.

408 Pennell, M.W., J.M. Eastman, G.J. Slater, J.W. Brown, J.C. Uyeda, R.G. FitzJohn, M.E. Alfaro &  
 409 L.J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting  
 410 macroevolutionary models to phylogenetic trees. *Bioinformatics*, **15**, 2216-2218.

411 Pigot, A.L. & Etienne, R.S. (2015). A new dynamic null model for phylogenetic  
 412 community structure. *Ecology Letters*, **18**, 153–163.

413 Rabosky, D.L. (2014). Automatic detection of key innovations, rate shifts, and diversity-  
 414 dependence on phylogenetic trees. *PloS One*, **9**, e89543.

415 Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and  
 416 other things). *Methods in Ecology and Evolution*, **3**, 217–223.

417 Rolland, J., Condamine, F.L., Jiguet, F. & Morlon, H. (2014). Faster Speciation and Reduced  
 418 Extinction in the Tropics Contribute to the Mammalian Latitudinal Diversity Gradient.  
 419 *PLoS Biology*, **12**, e1001775.  
 420 Silvestro, D., Schnitzler, J. & Zizka, G. A Bayesian framework to estimate diversification  
 421 rates and their variation through time and space. *BMC Evolutionary Biology*, **11**, 311.  
 422 Soetaert, K.E.R., Petzoldt, T. & Setzer, R.W. (2010). Solving differential equations in R:  
 423 package deSolve. *Journal of Statistical Software*, **33**.  
 424 Stadler, T. (2011a). Mammalian phylogeny reveals recent diversification rate shifts.  
 425 *Proceedings of the National Academy of Sciences of the United States of America*, **108**,  
 426 6187–6192.  
 427 Stadler, T. (2011b). Simulating Trees with a Fixed Number of Extant Species. *Systematic*  
 428 *Biology*, **60**, 676–684.  
 429 Zachos, J.C., Dickens, G.R. & Zeebe, R.E. (2008). An early Cenozoic perspective on  
 430 greenhouse warming and carbon-cycle dynamics. *Nature*, **451**, 279–283.  
 431  
 432  
 433  
 434

435 **Table 1. Major functions available in RPANDA**

Function	Description
<b>Characterising and comparing phylogenies using spectral densities</b>	
BICcompare	computes BIC values assessing the support of modalities in a phylogeny
JSDtree	computes the Jensen-Shannon distance between phylogenies
JSDtree_cluster	clusters phylogenies using hierarchical and k-medoids clustering
spectR	computes the eigenvalues of a phylogeny, and returns the principal eigenvalue, the skewness, and kurtosis of the spectral density profile, and the eigengap
plot_BICcompare	plots modalities on a phylogenetic tree
plot_spectR	plots the spectral density and eigenvalues of a phylogeny ranked in descending order
<b>Fitting models of diversification to phylogenies</b>	
fit_bd	fits a birth-death model to a phylogeny
fit_coal_cst	fits an equilibrium model with constant diversity through time to a phylogeny, using the coalescent model
fit_coal_var	fits a model with expanding diversity through time to a phylogeny, using the coalescent model
fit_env	fits a birth-death model with environmental dependency to a phylogeny

<code>fit_sgd</code>	fits the model of Speciation by Genetic Differentiation to a phylogeny
<code>likelihood_bd</code>	computes the likelihood corresponding to the birth-death model
<code>likelihood_coal_cst</code>	computes the likelihood corresponding to the coalescent model with constant diversity through time
<code>likelihood_coal_var</code>	computes the likelihood corresponding to the coalescent model with expanding diversity through time
<code>likelihood_sgd</code>	computes the likelihood corresponding to the model of Speciation by Genetic Differentiation
<code>plot_fit_bd</code>	plots diversification rates through time curves
<code>plot_fit_env</code>	plots diversification rates as a function of one or several environmental variables and time
<code>plot_dtt</code>	computes and plots diversity through time curves
<code>sim_sgd</code>	simulates a phylogeny arising from the model of Speciation by Genetic Differentiation

436

437

**Figure 1** Spectral density plot of the Phyllostomidae (left panel) and corresponding eigenvalues ranked in descending order (right panel). There is a clear gap between the third and fourth eigenvalue (indicated by an arrow), suggesting three modes of division in the phylogeny.

**Figure 2** Phyllostomidae phylogeny with branches coloured according to the 3 branching patterns (or modalities) identified by the eigengap, as given by k-medoid clustering.

**Figure 3** Heatmap and hierarchical cluster showing the pairwise similarities between the 25 Phyllostomidae genera with more than one species.

**Figure 4** Plots showing the estimated a) speciation, b) extinction, and c) net diversification rates through time for the Phyllostomidae phylogeny, output of the `plot_fit_bd` function.

**Figure 5** Plot showing the estimated accumulation of species richness through time for the Phyllostomidae phylogeny, output of the `plot_dtt` function.

**Figure 6** Plots showing the estimated speciation (a,b) and net diversification (c,d) rates as a function of time (a,c) and temperature (b,d) for the Phyllostomidae phylogeny, output of the `plot_fit_env` function.